Selection-Based Versus Topography-Based Responding: An Important Distinction for Stimulus Equivalence?

David A. D. Polson and Joseph A. Parsons
University of Victoria

English-speaking subjects were taught 16 English-French word pairs. Within any given trial, one word from each pair was presented; for eight items, subjects were to select its counterpart from an array of words in the other language (selection-based training), and for the other eight items there was no array and subjects were to type its counterpart (topography-based training). In Experiment 1, all items were trained from French to English, and later, interspersed no-feedback probe trials tested for the emergence of the reversed relations. Half of the eight selection-trained items were tested in the selection-based mode and half were tested in the topography-based mode; similarly, half of the eight topography-trained items were tested in the selection-based mode and half were tested in the topography-based mode. On the first reversal test trial, all 7 subjects scored 100% correct for the selection-tested items; in contrast, 5 of the 7 subjects scored 0% or near 0% correct for the topography-tested items, which improved to varying degrees with repeated testing. The training response mode affected neither acquisition rate nor reversal test trial performance. In Experiment 2, all items were tested in the topography-based mode only, and subjects were exposed to nine consecutive reversal test trials prior to interspersed probe testing. Improved accuracy across reversal test trials was not observed until the conditions of probe testing were instated, an indication that reexposure to the trained relations was a crucial component of delayed emergence. In Experiment 3, all items were trained from English to French, resulting in subjects typing a familiar rather than an unfamiliar word on reversal test trials. Accuracy on reversal test trials was considerably better than in the previous two experiments. We discuss the implications of our findings for stimulus equivalence research.

Michael's (1985) distinction between two types of behavior, selection-based responding and topography-based responding, may be relevant to the generality of the stimulus equivalence phenomenon. Selection-based responding involves conditional relations, whereas topography-based responding involves discriminative relations (Hall & Chase, 1991). In selection-based responding, "the unit of verbal behavior can be described as an increased control of the pointing response by a particular stimulus as the result of the presence of a different stimulus (or the strength of a particular establishing operation)" (Michael, 1985, p. 2). For example, a child, when presented with a picture of a giraffe, becomes more likely to point to the word *giraffe* on a page of text containing several animal names; and, when presented with a picture of an elephant, becomes more likely to point to the word *elephant.* In both instances, the form of the response (pointing) is identical, and which printed-word stimulus controls that response is conditional upon the presence of yet another stimulus (a picture). In topography-based

Portions of the data were presented at the 1996 convention of the Northwest Association for Behavior Analysis in Ellensburg, Washington, the 1995 convention of the Northwest Association for Behavior Analysis in Eugene, Oregon, the 1995 convention of the Association for Behavior Analysis in Washington, D.C., and the 1994 convention of the Northwest Association for Behavior Analysis in Seattle, Washington. Judy Bredin and Kezia Tan ran subjects in Experiment 1, and Judy Bredin, Maurene Bedich, and Kim Hanson ran subjects in Experiments 2 and 3, as part of an undergraduate research course. We thank these students, as well as Diana Grabavac, for their contributions.

Requests for reprints should be sent to David Polson, Department of Psychology, University of Victoria, P.O. Box 3050, Victoria, British Columbia V8W 3P5, Canada (E-mail: dpolson@uvic.ca).

responding, "the unit of verbal behavior can be described as an increased strength of distinguishable topography given some specific controlling variable" (Michael, 1985, p. 2). Here, the presence of each animal picture would evoke a response of distinguishable form: Depending on which picture was presented, the child might say "giraffe" or "elephant."

The match-to-sample (MTS) procedure employed by equivalence researchers (e.g., Sidman, 1986; Sidman & Tailby, 1982) trains and tests selection-based responding. For example, after training the "see picture/select word" relation described above, the task might be reversed: Each word could be presented along with an array of animal pictures, and the new task becomes "see word/select picture." Symmetry would be demonstrated if the child chooses the picture that goes with the word without ever having received reinforcement for doing so (an *emergent relation*; Sidman, 1986).

According to Hall and Chase (1991), the mathematical definition of equivalence, originally applied to conditional discrimination performances by Sidman and Tailby (1982), does not restrict equivalence research to a selection-based paradigm. Hall and Chase argue that a subject can demonstrate equivalence not only by emitting topographically indistinguishable responses to *select* stimuli, but also by emitting topographically distinguishable responses to *produce* stimuli. Regarding the latter, suppose we train a relation between two stimuli, "long neck" and "giraffe," by reinforcing the *saying* of "long neck" whenever we present the vocal stimulus "giraffe" (hear A/say B, an intraverbal relation; Skinner, 1957). To test for symmetry, we would present the vocal stimulus "long neck" and record whether or not the *saying* of "giraffe" (hear B/say A) emerges. Hall and Chase (p. 116) point out that to be consistent with the mathematical definition of equivalence, the nature of the task (Relation R) must remain the same for the trained and tested relations. Thus, a case in which the child receives reinforcement for *selecting* the picture of the giraffe upon hearing "giraffe" (hear A/select B), and later can *say* "giraffe" in the presence of the picture of the giraffe (see B/say A) with no additional training, would not constitute an instance of symmetry, because R differs during training (selecting upon hearing) and testing (saying upon seeing).

The present article is concerned with only one of the three defining relations of equivalence: symmetry (the other two being reflexivity and transitivity; Sidman & Tailby, 1982). Consistent with Hall and Chase (1991), we distinguish between two types (see also Polson, Grabavac, & Parsons, 1997). In *selection-based symmetry*, the relation "in the presence of Stimulus A, select Stimulus B" is reinforced, and the relation "in the presence of Stimulus B, select Stimulus A" emerges. In *topography-based symmetry*, the relation "in the presence of Stimulus A, produce Stimulus B" is reinforced, and the relation "in the presence of Stimulus B, produce Stimulus A" emerges, for which the task (e.g., saying upon hearing) is the same for both the trained and emergent relations.

Although proof for selection-based symmetry in human subjects is extensive (see, e.g., Sidman, 1994), its topography-based counterpart remains relatively unexplored. One exception is a study by Polson et al. (1997). College students, who were good typists, were taught 16 English-French word pairs (intraverbals) by requiring them to type one word of the pair when presented with the other word as a textual stimulus.[1] In Phase 1, half of the intraver-

[1] Poor typists were excluded, the reasoning being that for each letter in a response word they are likely to point to and press that character only after scanning the keyboard and encountering it as a visual stimulus (selection-based responding). Good typists, however, do not require an array to scan; rather, they string together letters fluently for each word, the total of which is a unit unto itself, one that is topographically dis-

bals were taught from French to English and half from English to French (see A/type B). Then, in Phase 2, training continued with the stimulus and response terms of each intraverbal reversed, that is, items trained from French to English in Phase 1 were trained from English to French in Phase 2, and items trained from English to French in Phase 1 were trained from French to English in Phase 2 (see B/type A). Feedback was provided throughout the experiment. Topography-based symmetry, as reflected by accuracy scores on the first trial in Phase 2, was generally poor: Eight of the 9 subjects averaged only 29% correct when asked to respond to the reversed relations for the first time. In addition, test performance was generally worse for items in which the response term was an unfamiliar (i.e., French) word. In contrast to these results, selection-based symmetry is often reported to occur at near-perfect levels when first tested with comparable subjects (i.e., college students, e.g., Dougher, Augustson, Markham, Greenway, & Wulfert, 1994; Lane & Critchfield, 1996; Mandell & Sheen, 1994), as well as with other, presumably less verbally sophisticated individuals, such as fifth- and sixth-grade students (Lynch & Cuvo, 1995), 5- to 7-year-old children (Sidman & Tailby, 1982), and institutionalized retarded young men (Sidman, Willson-Morris, & Kirk, 1986).

The findings by Polson et al. (1997), however, are consistent with symmetry research in the paired-associates literature (e.g., Feldman & Underwood, 1957; Jantz & Underwood, 1958; Nelson, 1972; Wollen & Allison, 1968). In these studies, subjects learn to vocalize one item of a pair when presented with the other as a textual stimulus (see A/say B). After reaching a specified mastery level or working through a predetermined number of trials, the order of the items is reversed (see B/say A). Should the see B/say A relation emerge without feedback, then this would be consistent with our definition of topography-based symmetry. One representative study was conducted by Jantz and Underwood (1958), who trained see A/say B items over either 4, 12, or 24 trials; the mean numbers correct on the first see B/say A test trial were 1.2, 3.3, and 4.0 out of eight, respectively. Nelson (1972), summarizing his decade-long research into symmetry using the paired-associates paradigm, concluded, "In our search for associative symmetry we have observed asymmetry almost everywhere we looked" (p. 150). Caution should be noted when drawing conclusions from this research, however, because the designs typically involved groups of subjects from which data were averaged and compared; individual data were ignored. Commenting on paired-associates and serial learning studies of equivalence, Sidman (1994) noted that false negatives may have been reported, because "group comparisons prevented the observation of individual successes" (p. 183). Inspection of single-subject data may have led to a different conclusion: Equivalence relations *could* form, even though they did not always do so. Indeed, in the study by Polson et al., performance on the symmetry test trial was outstanding for 1 subject relative to the 8 other subjects.

The lack of topography-based symmetry research in the equivalence literature is surprising given that stimulus equivalence has come to form the basis of sophisticated models for classifying behavioral units (Sidman, 1986) and interpreting complex human functioning such as verbal behavior (Hayes & Hayes, 1992). Certainly not all behavior is selection based (Moerk, 1990; Skinner, 1957). Not surprisingly, there have been recent criticisms of the selection-based MTS paradigm (Horne & Lowe, 1996) and calls for alternative methodologies for studying emergent-like behavioral phenomena (Saunders

tinct from the action involved in typing other words (topography-based responding). Polson et al. (1997) discuss this issue in greater detail.

& Green, 1996). As Michael (1985) explained, from a behavioral perspective, selection-based responding and topography-based responding contain idiosyncrasies that could affect ease of acquisition, control by motivational variables, and resistance to disruption (see also Shafer, 1993). For example, selection-based responding consists of an extra degree of conditionality and requires an effective scanning repertoire, two possible impediments to learning; topography-based responding involves point-to-point correspondence between response form and response product (Skinner, 1957), a factor that might facilitate learning. Numerous prerequisite behaviors unique to selection-based responding have also been noted by equivalence researchers (Sidman, 1994, chap. 3). Applied research with persons with developmental disabilities supports Michael's hypothesis in showing differential performance resulting from selection-based training and topography-based training of beginning verbal repertoires (Sundberg & Sundberg, 1990; Wraikat, Sundberg, & Michael, 1991).

Topography-based responding has played a role in equivalence research examining reversed relations, although it has not been identified as such. A number of studies documented that when a relation was trained in the selection-based mode (e.g., hear A/select B), some subjects were then able to appropriately respond to the reversed relation in the topography-based mode (see B/say A) without direct training (e.g., Lipkens, Hayes, & Hayes, 1993; Sidman, 1971; Sidman & Cresson, 1973; Sidman, Cresson, & Willson-Morris, 1974; Sidman & Tailby, 1982). Lipkens et al. (1993) also reported that after training relations in the topography-based mode (see A/say B) their subject was then able to appropriately respond to the reversed relations in the selection-based mode (hear B/select A) without reinforcement for doing so. The latter result is particularly impressive given that the subject was a normally developing infant (between 16 months and 27 months old). Other areas of research, employing different terminology, support these findings. Cuvo and Riva (1980), who studied "production" (see A/say B) and "comprehension" (hear B/select A), showed that training in either response mode resulted in substantial improvements in the other for both retarded and nonretarded children. Evidence of cross-modal transfer between reversed topography-based repertoires can also be found in the literature. Lee (1981) documented improvements in an unreinforced nonverbal repertoire (hear B/arrange A) following verbal training (see A/say B), and vice versa, for retarded youths. Similarly, Lee and Pegler (1982) demonstrated that overlearning in reading instruction (see A/say B) resulted in perfect spelling (hear B/spell A) without spelling instruction for 2 11-year-old children who could not read. In all these studies, because the nature of the task differed between the training and testing conditions, these emergent performances are probably best not classified as symmetry (Hall & Chase, 1991; Lowe & Horne, 1996, p. 327; Sidman, 1994, p. 227); nevertheless, because they demonstrate the establishment of repertoires without direct reinforcement, more detailed analyses of these phenomena are important for both theoretical and practical purposes (Alessi, 1987; Hayes & Hayes, 1992; Horne & Lowe, 1996; Sidman, 1994).

The series of experiments in this article began as a follow-up to Polson et al. (1997), but using a procedure more akin to that typically reported in equivalence research. As noted earlier, there was a discrepancy between their outcome for topography-based symmetry and the standard outcome for selection-based symmetry in the equivalence literature. However, any number of procedural variables irrelevant to the selection-based versus topography-based distinction might account for that discrepancy, such as stimulus materials, reinforcement contingencies, instructions, and so on. In Experiment 1, we directly compared topography-based symmetry and selection-based symmetry (and two variations thereof) within subjects, keeping constant such potentially confounding procedural variables. In doing so, we hoped to add to the literature seeking to clarify the nature of selection-based and topography-based verbal behavior (Potter & Brown, 1997). Then, building upon these results—indicating highly reliable symmetry only when the relations were tested in the selection-based mode—we restricted our focus to testing in the topography-based mode, and explored the necessary conditions for delayed emergence (Experiments 2 and 3) and the effects of familiarity (French vs. English words) of the stimulus and response terms (Experiment 3).

EXPERIMENT 1

In Experiment 1, we taught subjects to respond to some French words by selecting their English counterparts from a list (see A/select B; selection-based responding) and to respond to other French words by typing their English counterparts (see A/type B; topography-based responding). Then, we reversed the stimulus and response words and tested half the selection-trained items in the selection-based mode (see B/select A) and the other half in the topography-based mode (see B/type A); similarly, half the topography-trained items were tested in the selection-based mode and the other half in the topography-based mode.

Method

Subjects. All 7 participants were undergraduate students from the University of Victoria. Subjects were either recruited by the researchers and paid $8.00 per hour or selected from the Introductory Psychology 100 subject pool and compensated with bonus points toward their final grade in the course. These incentives were provided independent of their performance in the experiment. Requests for subjects asked for persons with little or no knowledge of the French language and a minimum typing speed of 30 words per minute (see Footnote 1). As an additional screening device, immediately prior to the experiment, all volunteers were required to complete a paper-and-pencil test in which they were asked to write the English equivalents of the French words that were to be used in the study. Only individuals who scored zero on this pretest served as subjects.

Table 1

The stimulus and response words in the training phase.

Stimulus	Response
bidon	flask
delit	crime
galet	wheel
juron	curse
malle	trunk
pitre	clown
regie	power
tison	brand
falot	light
colle	paste
honte	shame
ladre	miser
nacre	pearl
quete	group
sonde	drill
veuve	widow

Setting, apparatus, and trial contingencies. The experiment was conducted in a sound-attenuating chamber. The work station consisted of a table and a few chairs. An MS-DOS personal computer and a monochrome (amber on black) monitor were situated on the table. The software, written in Microsoft Quick Basic®, presented instructions, item displays, and feedback, and recorded stimulus and response events and their time of occurrence to the nearest 0.05 s.

The 16 French–English word pairs used throughout the experiment are presented in Table 1. The French and English words consisted of five letters, and the first letter of each French word

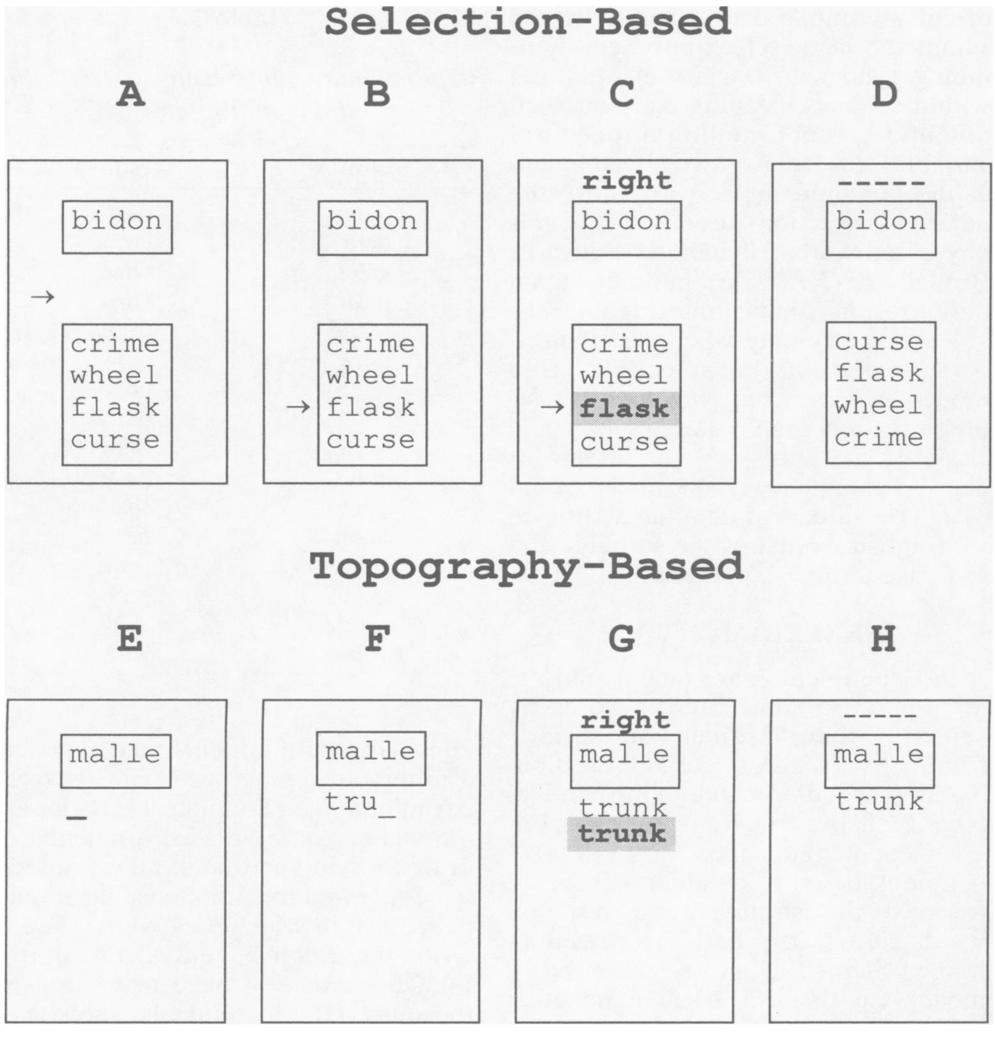

Fig. 1. Representations of the monitor display for an item presented in the selection-based mode and an item presented in the topography-based mode. Included are what a subject saw when the item first appeared in any given trial (Panel A and Panel E), midway through a response (Panel B and Panel F), after registering a correct response on a feedback trial (Panel C and Panel G), and after registering a response on a no-feedback trial (Panel D and Panel F).

differed from the first letter of all other French words.

A trial consisted of a complete pass through all 16 items. The order of item presentation was randomized from trial to trial. For each item, the computer program presented the stimulus word in a centered frame near the top of the computer screen. In the selection-based mode, a framed vertical array was centered three lines below the sample stimulus word (see Figure 1, Panel A). The array contained four alternative (comparison) response words; the incorrect choices were the response words from the three other items assigned to the same training/testing condition (to be elaborated in the *Independent Variable* section below). Within the array, the comparison response words were listed on consecutive lines in random order. A highlighted arrow appeared on the line above the array. The first press of the space bar moved the arrow down a line to the immediate left of the topmost response word. Pressing any other key before the first space-bar press produced only a low beep (50 Hz, 0.11 s).[2] This contingency remained in effect after the first space-bar press, with the exception of pressing the enter key. Each subsequent space-bar press moved the arrow down one line (see Figure 1, Panel B). Pressing the space bar when the arrow was beside the bottom response word returned the arrow to the topmost response word in a closed loop. A response word was selected by pressing the enter key when the arrow was beside it.

In the topography-based mode, the stimulus word was presented as described above; however, instead of the array, a highlighted and nonflashing cursor appeared on the line beneath the stimulus word (see Figure 1, Panel E). If the enter key was pressed first, there was no consequence. Typing a character displayed that character highlighted at the cursor and advanced the cursor (see Figure 1, Panel F). The back-space key was functional. Whenever the space bar or the tab key was pressed, the sole consequence was the low beep. Pressing the enter key at any point after typing at least one character registered the displayed string of characters to the left of the cursor as the response to the stimulus word.

After a response was entered in either mode, the consequences varied depending on the trial type, the response mode, and whether or not the response was correct. On feedback trials in both response modes, correct responses resulted in a high (rising 850 to 1600 Hz for 1.65 s) beep and the word "right" displayed in highlighted reversed video text centered on the line above the stimulus word (see Figure 1, Panel C and Panel G); incorrect responses resulted in a low (falling 800 to 50 Hz for 1.65 s) beep and the word "WRONG" displayed. In addition, in the selection-based mode, the correct response word in the array was switched to reversed video text to draw attention to it; in the topography-based mode, the correct response was presented in reversed video text on the line below the typed response. On no-feedback trials in both response modes, the sole programmed consequence of any registered response was "—" appearing in highlighted reversed video text centered above the stimulus word (see Figure 1, Panel D and Panel H). Pressing the enter key a second time in both modes brought forth the next item.

Throughout the experiment, there was no feedback for responses on every third trial; otherwise feedback was provided. There were no breaks between trials; items continued to be presented until the end of the session, at which time instructions appeared on the screen to "Please wait for the experimenter." If a session timed out while the subject was in the middle of a trial, this instruction was delayed until the end of that trial.

Procedure. To begin each session, the subject was seated in front of the computer and was required to read the instructions displayed on the monitor before proceeding (see the Appendix for a transcript of all instructions).

The experiment began with a 5-min demonstration session that introduced the subject to the experiment and the task. The experimenter was present throughout this session to answer questions. Only the English words from Table 1 were used for the demonstration task. For half of the English words, the subject was required to select the identical English word to the one presented; for the other half, the subject was required to type the identical English word to the one presented. The response mode (selection based or topography based) for each English word was the same response mode assigned

[2] This immediate auditory feedback was provided to cue subjects to the required response mode. Pilot testing revealed that, despite the visual display, subjects sometimes attempted to type the response word in the selection-based mode and to press the space bar in the topography-based mode.

Table 2

Percentage of correct responses for the baseline relations on feedback and no-feedback trials across all sessions in the training phase and the probe test phase.

Sessions	S1	S2	S3	S4
Training phase				
1. Feedback trials	77.1 (9)	75.0 (10)	38.5 (6)	51.9 (10)
No-feedback trials	87.5 (4)	85.0 (5)	56.2 (3)	57.5 (5)
2. Feedback trials	100 (13)	97.3 (14)	86.1 (9)	88.9 (13)
No-feedback trials	100 (6)	99.1 (7)	85.9 (4)	93.7 (6)
3. Feedback trials			97.9 (12)	96.1 (16)
No-feedback trials			100 (5)	99.1 (7)
4. Feedback trials			97.6 (13)	
No-feedback trials			100 (6)	
5. Feedback trials			98.6 (13)	
No-feedback trials			97.9 (6)	
Probe test phase				
1. Feedback trials	99.5 (12)	97.3 (14)	98.4 (10)	98.6 (13)
No-feedback trials	100 (3)	100 (3)	100 (2)	97.9 (3)
2. Feedback trials	99.5 (13)	98.7 (14)	98.1 (10)	100 (12)
No-feedback trials	97.9 (3)	98.4 (4)	100 (3)	93.7 (3)

Note. Parentheses indicate number of trials. One trial equals a complete run through the 16-item list in Table 1.

to that word in the subsequent training phase. After the 5-min demonstration, the experimenter left the room and a 15-min pretraining session ensued that employed this same task. Pretraining was included to function as a warm-up, in which subjects learned to respond both quickly and accurately to a mixture of items in the selection-based mode and the topography-based mode.

The training phase came next, consisting of multiple 15-min sessions. In each training session, the French words from Table 1 were presented, and the subject was required to select (see A/select B) or type (see A/type B) the English counterparts, depending on the training response mode assigned to each item. These 16 trained relations will be referred to as the baseline relations. As shown in Table 2, the number of sessions in the training phase varied: Training was terminated for some subjects soon after stable near-perfect performance within a session was achieved (e.g., S1), whereas for other subjects training was extended for a number of sessions beyond that point (e.g., S3). Recall that every third trial throughout the experiment was a no-feedback trial. No-feedback trials were included during the training phase to prepare subjects for the upcoming no-feedback reversal test trials.

The training phase was followed by the probe test phase, consisting of one or more probe test sessions. Baseline relations continued to be presented as before, but now, to probe for the emergence of the reversed relations, a reversal test trial was programmed every sixth trial. On these probe trials, the stimulus and response words for all 16 pairs were interchanged (i.e., the English, rather than the French, words served as the stimuli and the French, rather than the English, words served as the correct responses). In addition, the response mode for half the selection-trained items and half the topography-trained items also changed to the other response mode, depending on the testing response mode assigned to each item (see B/select A or see B/type A).

Table 2

Extended.

S5	S6	S7
33.0 (7)	35.2 (8)	52.3 (8)
41.7 (3)	43.7 (3)	75.0 (4)
85.9 (8)	82.5 (10)	98.4 (12)
90.6 (4)	88.7 (5)	100 (5)
99.0 (12)	97.8 (14)	98.7 (14)
100 (6)	99.0 (6)	97.9 (6)
	95.5 (14)	
	92.9 (7)	
	94.7 (14)	
	97.3 (7)	
100 (12)	96.3 (12)	97.9 (12)
97.9 (3)	95.8 (3)	100 (3)
97.9 (12)	97.4 (12)	97.9 (12)
100 (3)	100 (3)	100 (3)

Because every third trial was a no-feedback trial, feedback was never provided on reversal test trials (i.e., Trials 6, 12, 18) and some baseline trials (Trials 3, 9, 15, etc.). No-feedback baseline trials were included in the probe test phase for comparison purposes, to rule out the absence of feedback on reversal test trials as a confounding variable.

With the exception of the 5-min demonstration session, the experimenter was absent from the chamber while the subject was engaged with the computer program. All 15-min sessions were separated by a short break (approximately 5 min), during which the subject left the chamber while the experimenter prepared the computer for the next session. The total duration of the experiment for each subject was generally between 2.5 to 3 hr.

Independent variable. There were four conditions, defined by the response mode in which the baseline relations were trained (selection trained or topography trained) and the response mode in which the reversed relations were tested (selection tested or topography tested). These four conditions can be summarized as (a) sel/sel (selection trained/selection tested), (b) sel/top, (c) top/sel, and (d) top/top. Which four of the 16 items from Table 1 were assigned to each of these four conditions was randomly determined for each subject.

Results and Discussion

Training phase. Table 2 shows the percentage of correct responses for the baseline relations within each session in the training phase on both feedback trials and no-feedback trials. With the exception of the first training session, accuracy did not differ as a function of feedback. Within the first training session, the scores were consistently higher on no-feedback trials, primarily because of the rapid learning that occurred early on and the fact that the very first two trials were feedback trials. Thus, the low scores on these first two trials reduced the overall mean accuracy for feedback trials relative to no-feedback trials in that session. By the end of the training phase, accuracy was near perfect: All subjects scored 95% correct or above on both feedback trials and no-feedback trials within the 15-min training session immediately preceding the probe test phase. To put this in perspective, if a subject completed 21 trials in that session (e.g., S2), then that subject responded 21 times to each of the 16 baseline relations, for a total of 336 items within approximately 15 min, being correct on at least 95% of those items.

To assess the effects of the training response mode on acquisition of the baseline relations, Figure 2 displays the number of trials to criterion for the eight selection-trained items and the eight topography-trained items across subjects, the criterion being three consecutive 100% correct trials. The reasons for this stringent criterion are twofold. First, if subjects were guessing at one or more of the answers in the selection-based mode, it is extremely unlikely that they would be able to do so

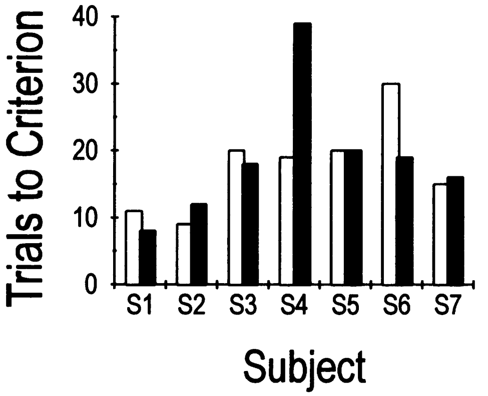

Fig. 2. Number of trials to criterion for the eight selection-trained items (white bars) and the eight topography-trained items (black bars) across all subjects in Experiment 1. Criterion was three consecutive 100% correct trials.

correctly over three consecutive trials. Second, a block of three trials necessarily includes both two feedback and one no-feedback trials. Figure 2 reveals that subjects varied as to whether the selection-trained items or the topography-trained items were acquired to criterion sooner.

Probe test phase. Table 2 also shows the percentage of correct responses for the baseline relations within each session in the probe test phase on both feedback trials and no-feedback trials. The near-perfect scores observed during later sessions in the training phase continued throughout the probe test phase. Thus, accuracy for the baseline relations was unaffected by the introduction of the interspersed reversal test trials.

Inspection of the raw data indicated that the performance on reversal test trials was differentially affected by how the items were tested and *not* by how they were trained; thus, the reversal test data are presented grouped together on the basis of the testing response mode. Figure 3 shows the percentage of correct responses on reversal test trials within each session for the eight selection-tested items and the eight topography-tested items. For the topography-tested items, a response was deemed correct only if all the characters typed and entered by the subject for a particular response word matched the actual spelling of that word as shown in Table 1 (whole-word accuracy). Considering only the first reversal test trial for selection-tested items, all 7 subjects scored 100% correct. In stark contrast, for topography-tested items, 3 subjects (S3, S5, and S6) scored 0% correct, 2 subjects (S4 and S7) scored near 0% correct, and the remaining 2 subjects scored 87.5% (S1) and 75% (S2) correct. Considering performance across remaining reversal test trials, all subjects maintained perfect or near-perfect scores for the selection-tested items, and they all demonstrated at least some improvement for the topography-tested items, ranging from very little (S6) to considerable (S4).

The gray bars in Figure 3 display accuracy for the first letter of the response word for the topography-tested items (first-letter accuracy). Accuracy was considerably better when calculated in this way, although scores of 100% correct were observed for only 2 subjects (S1 and S2) on the first reversal test trial and 2 other subjects (S4 and S7) on later reversal test trials. In all cases, whenever performance was below 100% correct for whole-word accuracy, first-letter accuracy was greater. Consistent with whole-word accuracy, first-letter accuracy gradually improved across reversal test trials for the 5 lower scoring subjects.

EXPERIMENT 2

Equivalence researchers have often demonstrated that conditional relations emerge with repeated nonreinforced testing (e.g., Lazar, Davis-Lang, & Sanchez, 1984; Sidman, Kirk, & Willson-Morris, 1985; Sidman et al., 1986), including symmetrical relations (Bush, Sidman, & de Rose, 1989; Gatch & Osborne, 1989; Lynch & Cuvo, 1995). This phenomenon has come to be known as *delayed emergence* (Sidman, 1994, pp. 273–278 and pp. 511–512). Although the perfect scores on the first

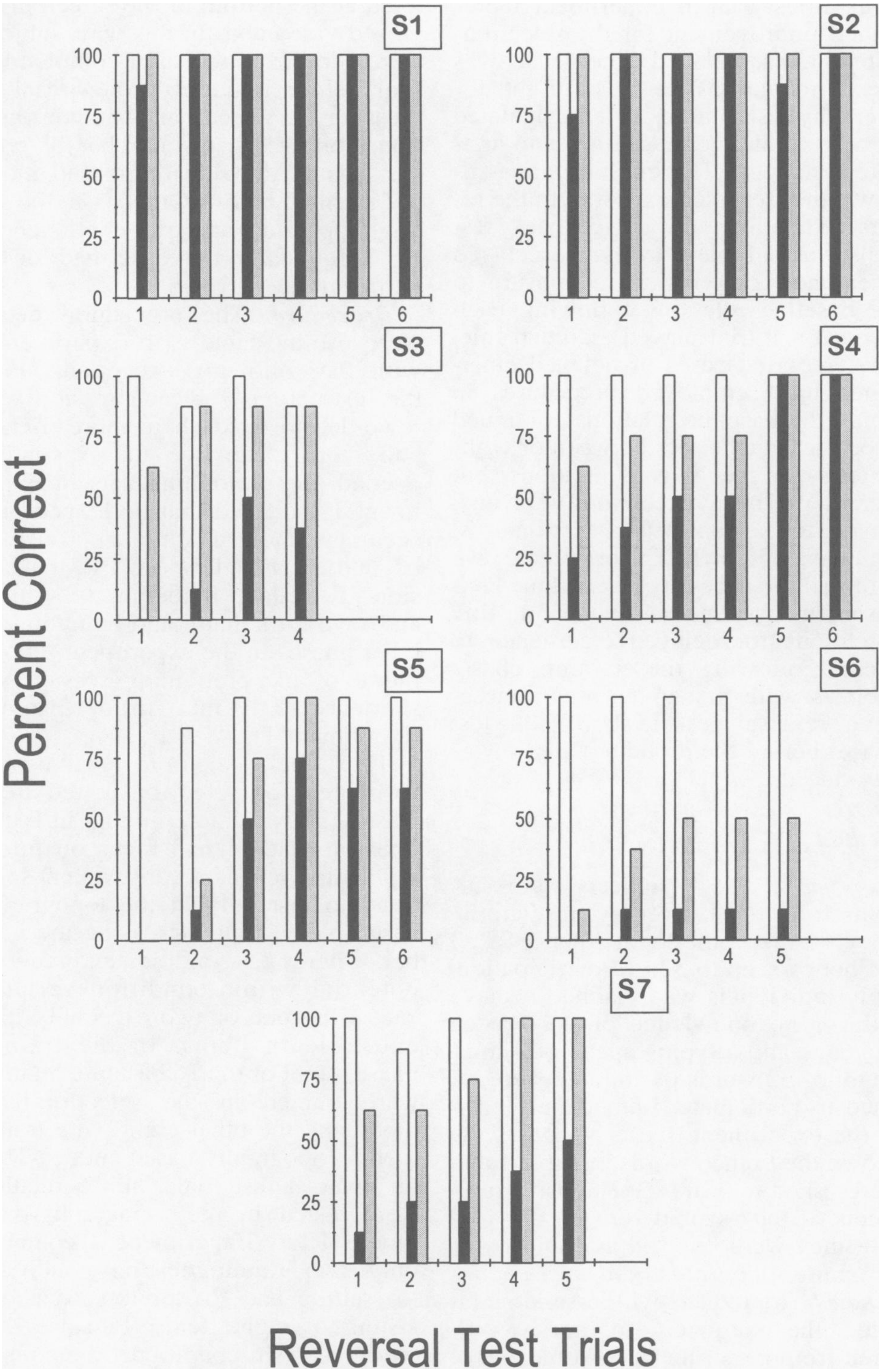

Fig. 3. Accuracy scores across reversal test trials for all subjects in Experiment 1. White bars represent the mean for the eight selection-tested items. Black bars represent the mean for the eight topography-tested items when determined by whole-word accuracy, and gray bars represent the mean for the same eight items when determined by first-letter accuracy.

reversal test trial in Experiment 1 precluded improvement for the selection-tested items, delayed emergence was shown to various degrees for the topography-tested items when calculated by both whole-word accuracy and first-letter accuracy. The question arises as to whether repeated exposure to the reversed relations on no-feedback test trials was sufficient to induce delayed emergence, or whether reexposure to the baseline relations following each reversal test trial played a crucial role. Most reported cases of delayed emergence have employed procedures in which the baseline relations continued to be presented either between or within tests for the derived relations, but there have been exceptions (Devany, Hayes, & Nelson, 1986; Markham & Dougher, 1993). In Experiment 2, we explored the necessity of continued exposure to baseline trials under testing conditions for delayed emergence to occur. Following the training phase, subjects were first given nine consecutive reversal test trials (block test phase) before the probe test phase was introduced.

Method

Subjects. The 5 subjects were students from the University of Victoria. They were promised and paid $8.00 per hour for up to 3 hr of participation. Only individuals who claimed to have little or no knowledge of the French language and a typing speed of a minimum of 30 words per minute were invited to participate. Immediately prior to the experiment, these persons were shown the French words in Table 1 and were asked to write the English equivalents. If they scored zero on this pretest, they were selected as subjects.

Setting, apparatus, and trial contingencies. To reduce within-session fatigue, the computer program was altered from Experiment 1. At the end of each trial, the screen cleared, the program paused, and the message "When you are ready, press space bar to continue" was displayed on two lines centered at the bottom of the screen in reversed video text. In this way, subjects were allowed to rest after completing a trial before initiating a new trial by pressing the space bar. Another variation, concerning the number of comparisons in the selection-based mode, is described below. Otherwise, the setting, apparatus, and trial contingencies were identical to that described for Experiment 1.

Procedure. The procedural details were also the same as in Experiment 1, with the following exceptions. First, the instructions for each session were reworded to make them more friendly and more clear (see the Appendix). Second, the pretraining data in Experiment 1 indicated that both speed and accuracy showed little improvement beyond the first few trials in that session. To reduce boredom by subjects and save time that could be allotted to other phases in the experiment, the duration of the pretraining session was shortened to 10 min (from 15 min in Experiment 1).

Third, the excellent reversal test performance for selection-tested items was a highly reliable finding in Experiment 1. Rather than focus on further replication in this regard, we chose instead to test only in the topography-based mode, thus increasing the number of items per condition and reducing potential within-condition variation caused by one or two especially difficult-to-learn items. In the training phase, eight of the 16 baseline relations were trained in the selection-based mode and the other eight were trained in the topography-based mode. Then, on reversal test trials, all 16 relations were tested in the topography-based mode. Thus, Experiment 2 contained only two training/testing conditions: (a) sel/top and (b) top/top, each consisting of eight items. Because there were only two conditions, the comparison array in the selection-based mode contained eight choices (rather than four, as in Experiment 1), the incorrect alternatives being the response words

Table 3

Percentage of correct responses for the baseline relations on feedback and no-feedback trials across all sessions in the training phase and the probe test phase.

Sessions	S8	S9	S10	S11	S12
Training phase					
1. Feedback trials	60.9 (8)	20.8 (6)	20.3 (8)	36.5 (6)	29.2 (6)
No-feedback trials	68.7 (4)	37.5 (2)	20.8 (3)	56.2 (3)	34.4 (2)
2. Feedback trials	89.6 (12)	70.8 (6)	49.1 (7)	89.4 (10)	73.2 (7)
No-feedback trials	90.0 (5)	83.3 (3)	52.1 (3)	95.3 (4)	75.0 (3)
3. Feedback trials	98.6 (13)	92.2 (8)	79.7 (8)	98.7 (14)	95.0 (10)
No-feedback trials	97.9 (6)	96.9 (4)	73.4 (4)	99.1 (7)	96.2 (5)
4. Feedback trials		97.5 (10)	92.4 (9)		100 (12)
No-feedback trials		96.9 (4)	90.6 (4)		100 (5)
Probe test phase					
1. Feedback trials	99.5 (12)	98.4 (12)	97.4 (12)	98.4 (12)	100 (12)
No-feedback trials	100 (3)	100 (3)	100 (3)	100 (3)	100 (3)
2. Feedback trials			100 (12)	99.5 (12)	99.0 (12)
No-feedback trials			97.9 (3)	97.9 (3)	100 (3)

Note. Parentheses indicate number of trials. One trial equals a complete run through the 16-item list in Table 1.

from the seven other items assigned to selection-based training.

Fourth, Experiment 2 included a block test phase immediately following the training phase. The block test phase

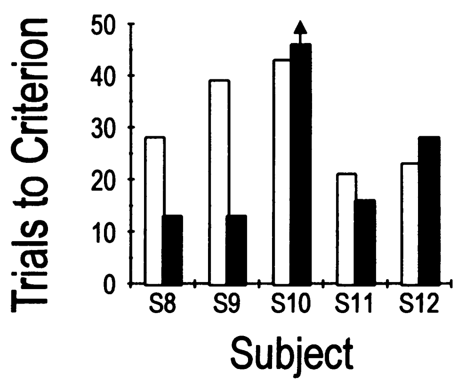

Fig. 4. Number of trials to criterion for the eight selection-trained items (white bars) and the eight topography-trained items (black bars) across all subjects in Experiment 2. Criterion was three consecutive 100% correct trials. The arrow above topography-trained items for S10 indicates that that subject did not achieve criterion within the 46 baseline trials that preceded the probe test phase.

consisted of a single session in which there were nine consecutive reversal test trials of the 16 relations. This was followed by the probe test phase.

Finally, to ensure that all subjects were exposed to the same number (i.e., three) of reversal test trials per probe test session, these sessions were terminated based upon the number of trials completed (18) rather than, as in Experiment 1, time elapsed (15 min).

Results and Discussion

Training phase. Table 3 presents the percentage of correct responses for the baseline relations within each session in the training phase on both feedback trials and no-feedback trials. Consistent with Experiment 1, accuracy did not differ as a function of feedback. During the final training session, near-perfect accuracy was observed for 4 of the 5 subjects, all of whom scored above 95% correct on both feedback trials and no-feedback trials; the exception (S10) scored above 90% correct on both trial types.

Figure 4 displays the number of trials to criterion for the eight selection-trained items and the eight topography-trained items across subjects, the criterion once again being three consecutive 100% correct trials. As observed in Experiment 1, subjects varied as to whether the selection-trained items or the topography-trained items were learned more rapidly. One subject (S10) never achieved criterion for the topography-trained items, although he did achieve two consecutive 100% correct trials at 45 trials, one of which was a no-feedback trial.

Block test phase. An inspection of the raw data once again revealed that performance on reversal test trials was unaffected by the training response mode. For ease of presentation, these data are presented independent of how the items were trained. Accuracy scores for the nine consecutive reversal test trials in the block test phase are shown to the left of the dotted line in Figure 5 for both whole-word accuracy and first-letter accuracy. On the first trial in the block test phase, low whole-word accuracy scores (44% correct or below) were observed for 4 of the 5 subjects; the exception (S8) obtained 75% correct. For all subjects, whole-word accuracy showed no signs of improvement across the eight remaining trials in the block test phase.

Accuracy scores were strikingly better when calculated for first-letter accuracy, even more so than the first-letter versus whole-word differences reported in Experiment 1. On the first trial in the block test phase, 4 of the 5 subjects (S8, S9, S11, and S12) scored 81% correct or above, and 1 of these subjects (S12) did not make any mistakes. There was no discernible upward trend across the eight remaining trials in the block test phase, although 1 subject (S11) improved slightly to eventually achieve 100% correct.

Probe test phase. Following the block test phase, interspersed training of the baseline relations was reinstated in the probe test phase. As observed in Experiment 1, the near-perfect accuracy scores for the baseline relations at the end of the training phase continued throughout the probe test phase (see Table 3). The subject (S10) who had performed below 95% correct on both feedback and no-feedback trials during the last training session improved to above 97% correct on both trial types during the first probe test session. Overall, the intervening block test phase as well as intermittent exposure to reversal test trials during the probe test phase did not disrupt near-perfect responding for the baseline relations.

Accuracy scores for the reversal test trials in the probe test phase (interspersed every sixth trial among baseline trials) are shown to the right of the dotted line in Figure 5 for both whole-word accuracy and first-letter accuracy. For all subjects, whole-word accuracy scores on the first reversal test trial in this phase increased from the last trial in the block test phase, with continued improvement over the remaining test trials. Consistent with Experiment 1, this trend was gradual. One subject (S11) eventually achieved 100% correct.

Perhaps due to a ceiling effect, the impact of introducing the probe test phase was less dramatic when considering first-letter accuracy as opposed to whole-word accuracy. Where improvement could occur, it generally did, and, by the end of the probe test phase, 2 of the 5 subjects (S11 and S12) had achieved scores of 100% correct, and the remaining 3 subjects (S8, S9, and S10) all scored as high as 94% correct.

EXPERIMENT 3

Polson et al. (1997) reported better topography-based symmetry for items trained from English to French than for items trained from French to English. Experiments 1 and 2 employed French-to-English training. In Experiment 3, the direction of training was switched, and the English, rather than the French, words functioned as the stimuli for the baseline relations. Should higher accuracy scores be observed on the reversal test trials for topography-tested

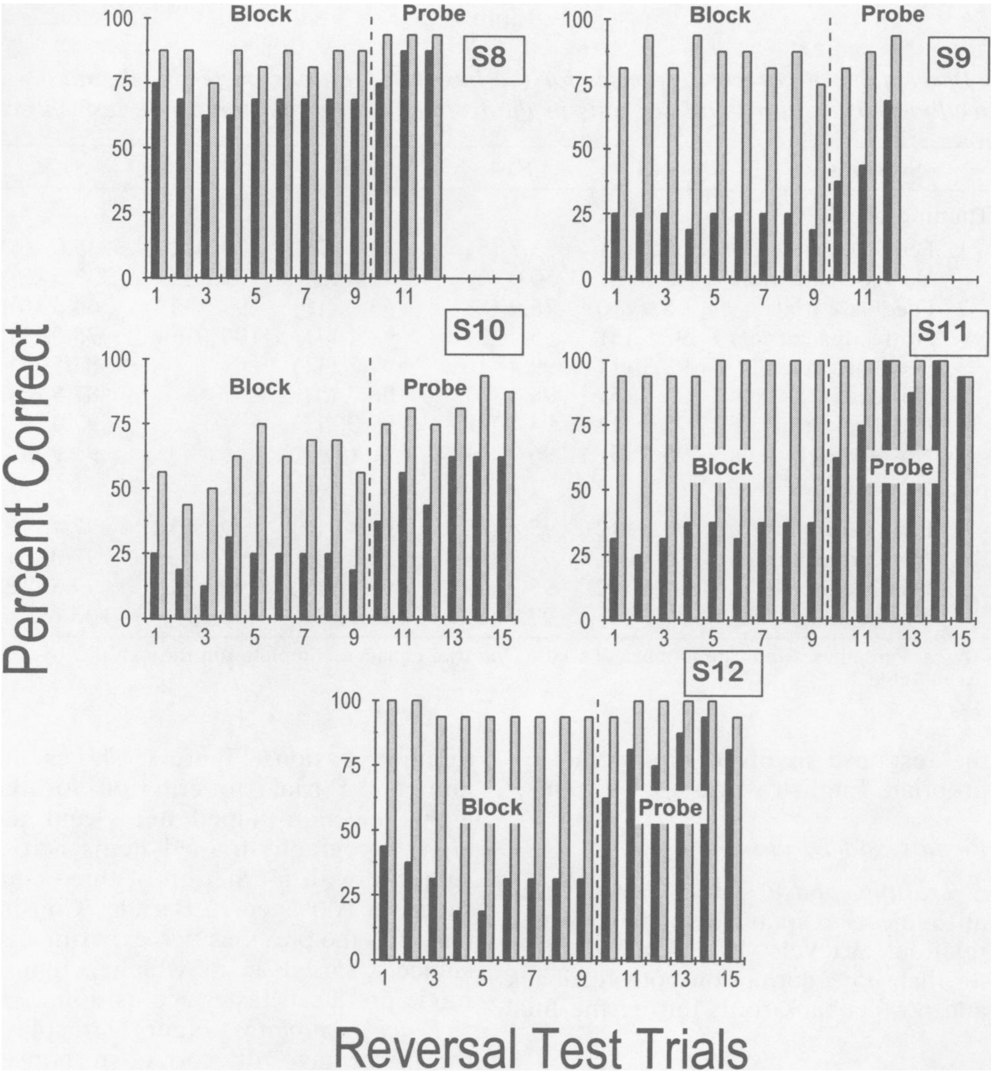

Fig. 5. Accuracy scores across reversal test trials for all subjects in Experiment 2. Black bars represent the mean for all 16 topography-tested items when determined by whole-word accuracy, and gray bars represent the mean for the same 16 topography-tested items when determined by first-letter accuracy.

items than in the previous two experiments, then this would provide a systematic replication of Polson et al.

Method

Five additional subjects participated in Experiment 3. All other details for subjects, setting, apparatus, trial contingencies, and procedure were identical to Experiment 2, with two exceptions. First, only the French words from Table 1 were used for the demonstration and pretraining tasks. Thus, subjects were required to select or type the French word identical to the one presented. Second, in the training phase, the English words were presented as stimuli and the response involved either typing or selecting the French counterparts to those words. Later, on reversal test trials, the French words were presented as stimuli and

Table 4

Percentage of correct responses for the baseline relations on feedback and no-feedback trials across all sessions in the training phase and the probe test phase.

Sessions	S13	S14	S15	S16	S17
Training phase					
1. Feedback trials	65.3 (8)	33.7 (5)	23.2 (7)	73.7 (10)	25.0 (6)
No-feedback trials	66.7 (3)	50.0 (2)	37.5 (3)	92.5 (5)	37.5 (2)
2. Feedback trials	96.9 (10)	78.9 (8)	93.1 (10)	99.5 (14)	64.6 (6)
No-feedback trials	97.5 (5)	85.4 (3)	98.4 (4)	100 (6)	75.0 (3)
3. Feedback trials	98.3 (11)	99.4 (10)	97.7 (11)		87.5 (7)
No-feedback trials	97.5 (5)	98.4 (4)	98.7 (5)		87.5 (3)
4. Feedback trials	98.3 (11)	94.4 (10)	99.0 (12)		97.9 (9)
No-feedback trials	96.2 (5)	96.9 (4)	100 (5)		96.9 (4)
Probe test phase					
1. Feedback trials	99.5 (12)	95.8 (12)	98.4 (12)	99.0 (12)	97.4 (12)
No-feedback trials	100 (3)	97.9 (3)	95.8 (3)	97.9 (3)	97.9 (3)
2. Feedback trials	99.5 (12)	95.3 (12)	99.5 (12)	99.5 (12)	99.5 (12)
No-feedback trials	97.9 (3)	97.9 (3)	93.7 (3)	100 (3)	100 (3)

Note. Parentheses indicate number of trials. One trial equals a complete run through the 16-item list in Table 1.

the response involved typing the appropriate English words for all items.

Results and Discussion

Training phase. Table 4 reveals that all subjects responded to the baseline relations at levels very close to or better than 95% correct on both feedback and no-feedback trials during the final training session. Figure 6 shows the number of trials to criterion for the eight selection-trained items and the eight topography-trained items across subjects, with a criterion of three consecutive 100% correct trials. Consistent with the previous two experiments, subjects varied as to which response mode produced faster acquisition.

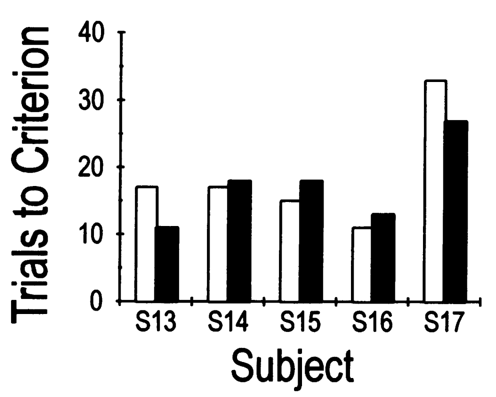

Fig. 6. Number of trials to criterion for the eight selection-trained items (white bars) and the eight topography-trained items (black bars) across all subjects in Experiment 3. Criterion was three consecutive 100% correct trials.

Block test phase. Figure 7 displays the percentage of correct responses when calculated by both whole-word accuracy and first-letter accuracy across all reversal test trials, beginning with the nine consecutive test trials in the block test phase. On the first of these trials, whole-word accuracy occurred at 75% correct or better for all subjects. Whole-word accuracy generally remained unchanged across the eight remaining trials in this phase, although 4 of the 5 subjects (S13, S14, S15, and S16) scored higher on at least one of these trials than on the first trial. Similar results are seen for first-letter accuracy. Although first-letter accuracy was sometimes higher than whole-word accuracy, the size of the differ-

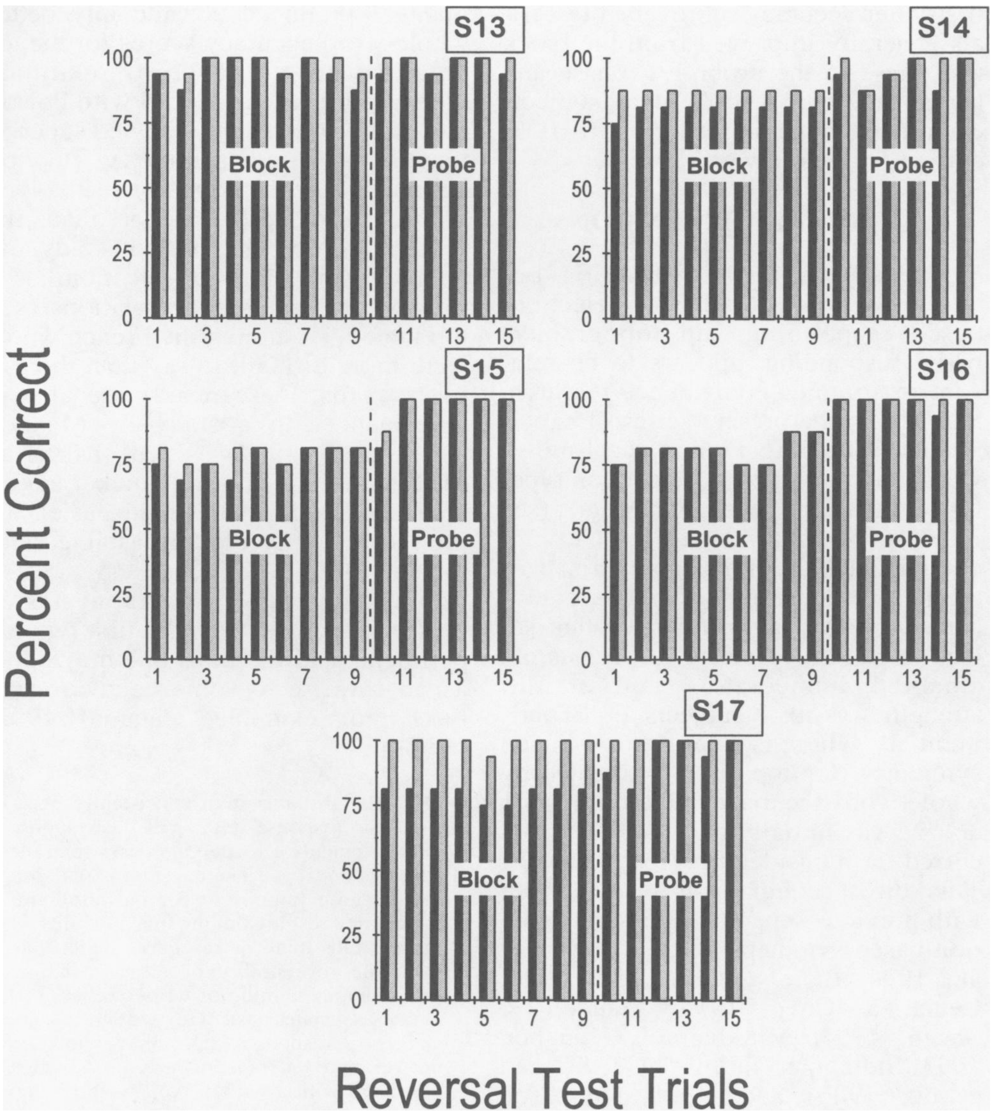

Fig. 7. Accuracy scores across reversal test trials for all subjects in Experiment 3. Black bars represent the mean for all 16 topography-tested items when determined by whole-word accuracy, and gray bars represent the mean for the same 16 topography-tested items when determined by first-letter accuracy.

ence and the consistency across subjects were not as great as observed in the previous two experiments when the tested response involved typing French words rather than English words. The most pronounced effect in this regard occurred for S17, who, with minor variations, consistently scored 100% correct for first-letter accuracy and 75% correct for whole-word accuracy throughout the phase.

Probe test phase. Table 4 shows that subjects responded to the reinstated baseline relations near perfectly on both feedback and no-feedback trials throughout the probe test phase. Figure 7 reveals that in most cases, whenever possible, whole-word accuracy and first-letter accuracy on reversal test trials generally improved from the block test phase to the probe test phase and across the probe test phase, all subjects eventually achieving scores of 100% correct for whole-word accuracy.

GENERAL DISCUSSION

Michael's (1985) distinction between two types of behavior, selection-based responding and topography-based responding, appears to be relevant to stimulus equivalence research. Subjects in the present series of experiments learned 16 French and English word relations (see A/select [or type] B), and later the reversed relations (see B/select [or type] A) were tested without feedback. A direct comparison of the two response modes in this equivalence-like paradigm revealed that selection-based symmetry (sel/sel items) emerged immediately and at full strength for all 7 subjects in Experiment 1, whereas topography-based symmetry (top/top items), as tested by whole-word accuracy in Experiments 1 and 2, was initially unreliable and occurred at considerably lower levels. This direct comparison is consistent with previous separate reports of selection-based symmetry (e.g., Dougher et al., 1994; Lane & Critchfield, 1996; Lynch & Cuvo, 1995; Mandell & Sheen, 1994; Markham & Dougher, 1993; Sidman & Tailby, 1982; Sidman et al., 1986) and topography-based symmetry (Polson et al., 1997). Inclusion of two other conditions (top/sel items in Experiment 1 and sel/top items in Experiments 1 and 2) revealed that the training response mode affected neither acquisition of the baseline relations nor reversal test trial performance.

In Experiments 1 and 2, the baseline relations were trained from French to English, and thus the reversed relations required typing French words. In Experiment 3, the direction of training was English to French, and now the reversed relations required typing the English words. This change in Experiment 3 produced considerably better whole-word accuracy scores for the topography-tested reversed relations. These results are in accord with Polson et al. (1997), who also showed superior topography-based symmetry (top/top items) when the reversal test involved typing the English rather than the French words. The present study extends this finding to selection-based training of the baseline relations (sel/top items). Assuming the French words were more difficult to say than the English words, these results are also in agreement with a study by Mandell and Sheen (1994) which, using an MTS paradigm, showed quicker equivalence class formation the easier the pronounceability of the baseline sample stimuli.

The discrepant topography-tested performances between the first two experiments and the third one may come as no surprise to some behavior analysts. For example, Catania (1992) writes,

> Once the stimulus item consistently occasions the response item, will the response item be equally effective in occasioning the stimulus item? . . . The question is not simple, because failures to demonstrate symmetry may result from the unavailability of the stimulus items as responses rather than from the reversibility of the association. For example, naming of written letters is a paired-associates task with written stimuli and vocal responses; if a child who does not yet write letters learns to say "A" when presented with a written A, it would be no surprise if the child could not write an A in response to the spoken letter. (pp. 287–288)

Like the child who has not yet learned to write the letter A in Catania's example, our subjects, who were screened on the basis of their ignorance of the French language, had presumably not yet learned to spell the French words. Their first requirement to do so in Experiments 1 and 2 occurred on the initial reversal test trial. Alternatively, the subjects in Experiment 3, who scored much higher on this trial, were typing English words that they had presumably produced in one form or another on many previous occasions. Thus, the initial topography-based test scores in the first two experiments may have reflected subjects' inability to spell the French words. Consistent with this suggestion is the fact that accuracy was considerably better on the first reversal test trial (although rarely perfect) when measured by the accuracy of typing the first letter of the French words. Thus, reversing a baseline relation often resulted in a subject producing part, but not all, of the baseline stimulus (an example of a similar phenomenon in the paired-associates literature is reported by Nelson, 1972, p. 135). This "part" happened to be a unique feature of that stimulus that set it apart from all others, given that every French word started with a different letter (see Table 1). Whether similar results would be obtained if some baseline stimuli had the same first letter, or whether other unique features of the baseline stimuli would then be produced, is a question that could be addressed by future research. Interestingly, paired-associates research suggests that the functional stimulus of low-meaning three-consonant trigrams is the first letter (Kausler, 1974, pp. 142–144). Combined with our data, it appears as if that which exerts stimulus control during training of the baseline relations is that which is most reliably produced during topography-based testing of the reversed relations. Future research might also consider examining the evocative effects of separate elements of the baseline stimuli to directly assess the correspondence between baseline stimulus control and the responses emitted on the topography-based test.

Although they are perhaps not surprising to many behavior analysts, our results should function as a caution to equivalence researchers when extrapolating from studies conducted using the MTS paradigm to real-life exemplars. Topography-based responding sometimes creeps into discussions of symmetry (e.g., Hayes & Hayes, 1992; Lipkens et al., 1993; Pierce & Epling, 1995; Wulfert & Hayes, 1988). Besides the thorny theoretical implications of ignoring response modality (Horne & Lowe, 1996, p. 232; Lowe & Horne, 1996, p. 327), our data reveal additional considerations. First, if the response component of the reversed relation already exists as part of a person's repertoire (Experiment 3), then better symmetry may be obtained than if that response topography has not yet been acquired as a unit (Experiments 1 and 2). In the latter case, initial tests of symmetry may reveal its complete absence (S3, S5, and S6) if the criterion for correct responses is the exact production of the baseline stimulus. To paraphrase Lee (1981), baseline training may not instate new topographies, but rather modify the stimulus control of existing topographies. Consistent with this speculation (and our results), she found that teaching moderately retarded children to tact "front" and "behind" relations (verbal responding) did not produce collateral improvements in responding to "front" and "behind" instructions (nonverbal responding) unless children demonstrated those instruction-following topographies prior to verbal training. Similarly, Lipkens et al. (1993) reported that see B/say A relations did not emerge following hear A/select B training until the child was taught to echo the vocal stimuli.

Our data suggest another concern with ignoring the selection-based versus topography-based distinction. Suppose that a unique property of a stimulus complex acquires control over responding during training of a baseline relation (Stromer, McIlvane, & Serna, 1993). In the MTS paradigm, symmetry testing may result in the subject *selecting* the comparison stimulus based only on that feature; nevertheless, this would be sufficient for a "correct" response. In a topography-based paradigm, symmetry testing may result in the subject *producing* a stimulus that contains only that feature; but, if an exact copy of the baseline stimulus is required, then the response would be scored as "incorrect." This has major implications for education. For example, a language student who studies for a test by looking at French words and writing their English equivalents would fail miserably should the test reverse those relations and require that he or she write the French words given their English counterparts, with no partial marks awarded for partially correct answers. The student may, however, do quite well should the test consist of multiple choice questions, even if the relations are reversed. It is not a far stretch to think of similar examples in other areas of education (e.g., science, math). In one related study, Polson, Wong, Parsons, and Grabavac (1991) trained college students to respond fluently to 20 behavioral definitions by typing the corresponding terms. When given paper-and-pencil tests that required them to write the definitions given the terms, they often provided some parts of the definitions, but their answers were rarely coherent.

The present study may contribute to a better understanding of delayed emergence. In Experiment 1, both whole-word accuracy and first-letter accuracy gradually increased across reversal test trials that were interspersed every sixth trial among baseline trials. In Experiment 2, no such increase was observed when reversal test trials were presented nine times consecutively; only later, with interspersed testing as in Experiment 1, was comparable improvement noted. (The high scores for first-letter accuracy in Experiment 2 and for both whole-word and first-letter accuracy in Experiment 3 impose a ceiling on this effect.) Our data suggest that reexposure to the baseline relations was crucial to delayed emergence. It is not unreasonable to suggest that following the first reversal test trial in Experiment 1 and the block test phase in Experiment 2, subjects did something special on subsequent baseline trials—something special that increased the likelihood that they would respond more accurately on the next reversal test trial. Perhaps the 3 best scoring subjects (S1, S2, and S8) in the first two experiments employed actions of this sort during the training phase, resulting in their superior performance on the first reversal test trial. If we can identify such precurrent behaviors (Parsons, Taylor, & Joyce, 1981; Polson & Parsons, 1994; Skinner, 1968), then we may have an additional tool for programming so-called "emergent" behavior rather than expecting it. This approach may help to explain why some subjects are immediately able to "relate" stimuli in tests of emergent performance and others are not, much in the same way that Blough (1959) illustrated why some pigeons are better able to match stimuli under delay conditions: Successful subjects emit adjunct behaviors that increase the probability of a "correct" response.

At this point we can only speculate as to what these precurrent behaviors might be. One possibility is that, for whatever reasons, subjects practiced or came to practice behaviors on baseline trials similar to those we deemed correct on reversal test trials (cf. Lee & Pegler, 1982). For example, on baseline trials in Experiments 1 and 2, a subject sees the French word stimulus and types its English counterpart; then, seeing that English word displayed on the screen, the subject attempts to spell covertly the French word without looking at it; finally, the subject compares his or her spelling to the displayed French word stimulus, resulting in automatic reinforcement if it is correct. This could help to explain why accuracy remained stable across the nine consecutive reversal test trials comprising the block test phase in Experiment 2: There was no opportunity for automatic reinforcement because the correct spelling of the French words was never displayed. However, there was opportunity in the subsequent probe test phase, when improved accuracy was noted, because the French words reappeared as stimuli on the baseline trials separating the reversal test trials. Interestingly, for any given reversal test trial in the probe test phase, the duration of the baseline trial following it was almost always longer than the duration of the baseline trial preceding it, suggesting that subjects may have slowed down on baseline trials following reversal test trials because they were now doing something extra.

Equivalence researchers have pointed out that although conditional discrimination training establishes the prerequisites for emergent relations, any given stimulus may belong to several classes (e.g., Devany et al., 1986; Sidman, 1994). Consequently, "many test trials may be required before one of the possibilities proves relevant on *every* test trial" and "the relation that is possible on every test trial will come eventually to provide the basis for choice" (Sidman, 1994, p. 275). If a relation indicative of equivalence is the only basis for consistent responding during testing, then we should see delayed emergence as other irrelevant stimulus properties lose control. Note that a number of our subjects were already responding based on a property that proved to be relevant on every test trial (i.e., the first letter) before the delayed emergence of a relation involving a more complex response property (i.e., the whole word) was observed. Whether this phenomenon of first producing part and then later all of a baseline stimulus is unique to topography-based symmetry or whether uncontrolled contextual variables (e.g., instructions) play a role is a matter for further research.

At the end of the probe test phase in Experiments 1 and 2, complete emergence of the see B/type A relation remained absent for many subjects even after exposure to as many as 15 reversal test trials. When calculated by whole-word accuracy, 6 of the 12 subjects had yet to score above 75% correct. When calculated by first-letter accuracy, 5 of the 12 subjects had yet to score 100% correct. One might argue that we had not yet established a context involving mutual entailment between the baseline relations and their reversed counterparts (cf. Hayes & Hayes, 1992). To do so, when the see B/type A relation failed to emerge, we could explicitly train it and, if whole-word accuracy is important, reinforce the correct spelling of the French words given the English words. Then, following the training of a new set of items (see C/type D), we might be more likely to see perfect whole-word accuracy scores when the reversed see D/type C relation is tested. In other words, with sufficient exemplars, complete emergence of the reversed relations on the very first test trial should be seen. Although this speculation awaits empirical support with respect to both the topography-based and the selection-based response modes, we counter that a symmetrical context for the word–word relations had been established in other ways. First, the expressed purpose of the experiment was to learn French and English word pairs, a point also implied in the instructions (see the Appendix). Second, in Experiment 1, subjects could presumably type anything for topography-tested items when shown the English words, but they were forced to chose a French word for selection-tested items, thus reversing the English and French word relations, on those very same test trials. Third, some subjects experienced as many as six cycles of training and reversal testing with the same set of items, in addition to the nine consecutive reversal test trials in the block test phase following the training phase. Finally, although the poorer performing subjects could still not accurately type all the French words by the end of the probe test phase, their ability to do so did gradually improve with repeated testing. All things considered, something more seemed to be involved than mastering one relation (see A/select [or type] B) and establishing a context in which its reversed topography-based counterpart (see B/type A) is important. We suggest closer analyses of prerequisite behaviors (e.g., the ability to produce the baseline stimulus) and of precurrent behaviors during baseline training (e.g., practicing the "emergent" behavior) that may facilitate or possibly even account for at least some performances indicative of stimulus equivalence.

REFERENCES

Alessi, G. (1987). Generative strategies and teaching for generalization. *The Analysis of Verbal Behavior, 5,* 15–27.

Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior, 2,* 151–160.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51,* 29–45.

Catania, A. C. (1992). *Learning* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Cuvo, A. J., & Riva, M. T. (1980). Generalization and transfer between comprehension and production: A comparison of retarded and nonretarded persons. *Journal of Applied Behavior Analysis, 13,* 315–331.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46,* 243–257.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62,* 331–351.

Feldman, S. M., & Underwood, B. J. (1957). Stimulus recall following paired associate learning. *Journal of Experimental Psychology, 53,* 11–15.

Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal of the Experimental Analysis of Behavior, 51,* 369–378.

Hall, G. A., & Chase, P. N. (1991). The relationship between stimulus equivalence and verbal behavior. *The Analysis of Verbal Behavior, 9,* 107–119.

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations and the evolution of behavior analysis. *American Psychologist, 47,* 1383–1395.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185–241.

Jantz, E. M., & Underwood, B. J. (1958). R-S learning as a function of meaningfulness and degree of S-R learning. *Journal of Experimental Psychology, 56,* 174–179.

Kausler, D. H. (1974). *Psychology of verbal learning and memory.* New York: Academic Press.

Lane, S. D., & Critchfield, T. S. (1996). Verbal self-reports of emergent relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior, 65,* 355–374.

Lazar, R. M., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in children. *Journal of the Experimental Analysis of Behavior, 41,* 251–266.

Lee, V. L. (1981). Prepositional phrases spoken and heard. *Journal of the Experimental Analysis of Behavior, 35,* 227–242.

Lee, V. L., & Pegler, A. M. (1982). Effects on spelling of training children to read. *Journal of the Experimental Analysis of Behavior, 37,* 311–322.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56,* 201–239.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 315–340.

Lynch, D. C., & Cuvo, A. J. (1995). Stimulus equivalence instruction of fraction-decimal relations. *Journal of Applied Behavior Analysis, 28,* 115–126.

Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Process, 32,* 29–46.

Markham, M. R., & Dougher, M. J. (1993). Compound stimuli in emergent stimulus relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 60,* 529–542.

Michael, J. (1985). Two kinds of verbal behavior plus a possible third. *The Analysis of Verbal Behavior, 3,* 2–5.

Moerk, E. L. (1990). Three-term contingency patterns in mother-child verbal interactions during first language acquisition. *Journal of the Experimental Analysis of Behavior, 54,* 293–305.

Nelson, S. E. (1972). In search of associative symmetry. In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood* (pp. 133–153). New York: Appleton-Century-Crofts.

Parsons, J. A., Taylor, D. C., & Joyce, T. M. (1981). Precurrent self-prompting operants in children: "Remembering." *Journal of* *the Experimental Analysis of Behavior, 36,* 253–266.

Pierce, W. D., & Epling, W. F. (1995). *Behavior analysis and learning*. Englewood Cliffs, NJ: Prentice Hall.

Polson, D. A. D., Grabavac, D., & Parsons, J. A. (1997). Intraverbal stimulus-response reversibility: Fluency, familiarity effects, and implications for stimulus equivalence. *The Analysis of Verbal Behavior, 14,* 19–40.

Polson, D. A. D., & Parsons, J. A. (1994). Precurrent contingencies: Behavior reinforced by altering reinforcement probability for other behavior. *Journal of the Experimental Analysis of Behavior, 61,* 427–439.

Polson, D. A. D., Wong, W. K., Parsons, J. A., & Grabavac, D. M. (1991, May). *Using Think Fast to examine variables affecting the efficacy of flash card instruction.* Poster presented at the 17th annual convention of the Association for Behavior Analysis, Atlanta.

Potter, B., & Brown, D. L. (1997). Review of studies examining the nature of selection-based and topography-based verbal behavior. *The Analysis of Verbal Behavior, 14,* 85–104.

Saunders, R. R., & Green, G. (1996). Naming and equivalence relations. *Journal of the Experimental Analysis of Behavior, 65,* 312–314.

Shafer, E. (1993). Teaching topography-based and selection-based verbal behavior to developmentally disabled individuals: Some considerations. *The Analysis of Verbal Behavior, 11,* 117–133.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative.

Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalencies in severe retardation. *American Journal of Mental Deficiency, 77,* 515–523.

Sidman, M., Cresson, O., Jr., & Willson-Morris, M. (1974). Acquisition of matching to sample via mediated transfer. *Journal of the Experimental Analysis of Behavior, 22,* 261–273.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43,* 21–42.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities, 6,* 1–19.

Skinner, B. F. (1957). *Verbal behavior.* Englewood Cliffs, NJ: Prentice Hall.

Skinner, B. F. (1968). *The technology of teaching*. Englewood Cliffs, NJ: Prentice Hall.

Stromer, R., McIlvane, W. J., & Serna, R. W. (1993). Complex stimulus control and equivalence. *The Psychological Record, 43,* 585–598.

Sundberg, C. T., & Sundberg, M. L. (1990). Comparing topography-based verbal behavior with stimulus selection-based verbal behavior. *The Analysis of Verbal Behavior, 8,* 31–41.

Wollen, K. A., & Allison, T. S. (1968). Latency and frequency of S-R and R-S associations as a function of recall time. *Journal of Verbal Learning and Verbal Behavior, 7,* 895–899.

Wraikat, R., Sundberg, C. T., & Michael, J. (1991). Topography-based and selection-based verbal behavior: A further comparison. *The Analysis of Verbal Behavior, 9,* 1–17.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50,* 125–144.

APPENDIX

Instructions for Experiment 1

Demonstration session. When you have read these instructions, please press the space bar to begin your instructional session. A stimulus word will appear in a box at the top of the screen. Your task involves two response forms. For the "fill-in-the-blank" items, simply type in the corresponding word underneath. You may backspace to correct errors. Press the enter key when your response is complete. For the "multiple choice" items, use the space bar to move the arrow to the corresponding word. Press the enter key when you have made your selection. Some of the items will include feedback and some will not. Continue with the task until the message "please wait for the experimenter" appears.

Pretraining session. This 15 minute session is the same as the last. Please work as QUICKLY and as ACCURATELY as possible. Continue until the message "please wait for the experimenter" appears.

First session of the training phase. Are you ready to learn some French words? This session is the same as the last in every respect except that the word stimuli are now French words. You are asked to select or type the corresponding ENGLISH word. It is expected that you will not initially know any of the words. Please remember to work as QUICKLY and as ACCURATELY as possible.

All remaining sessions. Carry on. . . . Remember to work quickly and accurately.

Instructions for Experiment 2 and Experiment 3

Demonstration session. Welcome! Here is what will happen during this experiment. A word will appear in a box at the top of the screen. Sometimes you will be prompted to type the correct word that goes along with it. Backspacing is permitted. To register your response, press the enter key. Other times you will be prompted to select the correct word that goes along with it from a list of alternatives. To do so, use the space bar to point an arrow to one of the alternatives, and then press the enter key to confirm your selection. For both question types, you may or may not be told if your response was correct. Pressing the enter key a second time clears the screen and presents the next word. When you come to the end of the list, you start again. For this first task, you are to simply type or select the same word as the one that appears at the top of the screen. Press the space bar to proceed.

Pretraining session. Your task is the same as it was in the previous session; however, the experimenter will not remain in the room to coach you. Please respond to the presented words as QUICKLY and ACCURATELY as possible. At the end of the session, the screen will read: "Wait for the experimenter." This session will last about 10 minutes. Press the space bar to proceed.

First session of the training phase. Now you will be working with English words and their French equivalents. Like before, a word will appear in a box at the top of the screen and, depending on the question type, you are to select or type the correct word that goes along with it. This time, however, the correct word is the equivalent word in the other language. In the beginning, you will not know the answers. Don't despair! With practice you will get better. Please remember to respond to the presented words as QUICKLY and ACCURATELY as possible. At the end of the session, the screen will read: "Wait for the experimenter." This session should last about 15 minutes. Press the space bar to proceed.

All remaining sessions. Carry on. Please remember to respond to the presented words as QUICKLY and ACCURATELY as possible. At the end of the session, the screen will read: "Wait for the experimenter." This session should last about 15 minutes. Press the space bar to proceed.